

Data Analysis and Interpretation Workshops

SPRING 2020

DATE	ITEM
Week 1	
1/21 - 1/25	
Week 2	<i>Intro Stata I (first of a 3 or 7 workshop series)</i>
1/26-2/1	Location and registration required for the 3 workshop series at http://bit.ly/stataSpring2020 OR for the 7 workshop series http://bit.ly/dataSeriesSpring2020
	Offered 3 times: Sun, Jan 26 4-6:30pm --- Mon, Jan 27 4:30-7pm --- Wed, Jan 29 4:30-7pm
	This introductory session is based on a survey conducted on workshop participants. We cover the basics of opening files, running commands, creating summary statistics and graphics, dealing with outliers, manipulating variables, and labelling variables. We also work through how to deal with errors (which are inevitable, but manageable, in any statistical package). This will be hands on, practical experience: each participant will be using the program in real time.
Week 3	<i>Introduction to R</i>
2/2-2/8	Location and registration TBA
	We cover the basics of opening files, running commands, creating summary statistics and graphics, dealing with outliers, and manipulating variables.. We also work through how to deal with errors (which are inevitable, but manageable, in any statistical package). This will be hands on, practical experience: each participant will be using the program in real time.
	<i>Introduction to Interpreting Regressions</i>
	Location and registration required at http://bit.ly/regressionsS20
	Offered 3 times: Mon, Feb 3 4:15-6:15pm --- Tue, Feb 4 12:45-2:45pm --- Wed, Feb 5 7-9pm
	Many books and papers present regression results to support their arguments. There are some simple ways to scan these tables that let you quickly understand the main points and common pitfalls. We cover how to read regression tables and graphs efficiently, what the numbers for coefficients and statistical significance mean, and common ways regressions are misused.
Week 4	<i>Manipulating and Analyzing Data (2nd of the 7 workshop series, or a standalone)</i>
2/9-2/15	Location and registration required at http://bit.ly/excelSpring2020
	Offered 2 times: Sun, Feb 9 4-6pm --- Mon, Feb 10 4:30-6:30pm
	Using a dataset on county demographics, presidential voting, and deaths of despair, this workshop covers (1) how to use Excel more effectively, (2) which tasks Excel can perform better than a statistical programming package such as R, SPSS, or Stata, and (3) how to organize data effectively.
Week 5	
2/16-2/22	
Week 6	<i>Manipulating and Analyzing Data in Stata (3rd of the 7 workshop series)</i>
2/23-3/1	Location and registration by attending prior Stata/Excel workshops or instructor permission
	Offered 2 times: Sun, Feb 23 4-6pm --- Mon, Feb 24 4:30-6:30pm
	This workshop covers (1) the importance of the unit of observation for understanding how to manipulate data, (2) merging multiple datasets, (3) reshaping data to change the unit of observation, and (4) collapsing data to a higher unit of observation.
Week 7	<i>Where to Find Data</i>
3/2-3/7	Location and registration TBA

Many research ideas are generated by creating questions answerable by existing data sources. We cover some common data repositories, navigating these sources, and searching for specific questions in surveys. We also create graphics in Google Sheets to summarize public opinion data.

break

Week 8 *Introduction to Stata II (2nd of the 3 workshop series, 4th of the 7 workshop series)*

3/15-3/21 Location and registration by attending Intro to Stata I or instructor permission

Offered thrice: Sun, March 22 4-6:30pm --- Mon, March 16 4:30-7pm --- Wed, March 18 4:30-7pm

Using a dataset on county demographics, presidential voting, and deaths of despair, this workshop covers several additional basics of data analysis. This includes scatterplots, boxplots, standard errors/standard deviations/the normal distribution, and the fundamentals of developing useful theories.

Week 9

3/22-3/28

Week 10 *Manipulating and Analyzing Data in Stata (5th of the 7 workshop series)*

3/29-4/4 Location and registration by attending prior Stata workshops or instructor permission

Mon, March 30 4:30-6:30pm (Also offered in Week 12)

This workshop builds on the prior workshops in the 7 workshop series. It covers the basics of loops, the importance of documenting your work and version control, and creating codebooks for your data. We also touch on more advanced techniques and tricks for processing data, including fuzzy matching, looping through files, systematically renaming variables, regular expressions, and invisible characters.

Week 11 *Introduction to Stata III (3rd of the 3 workshop series, 6th of the 7 workshop series)*

4/5-4/11 Location and registration by attending Intro to Stata II or instructor permission

Offered 3 times: Sun, April 5 4-6:30pm --- Mon, April 6 4:30-7pm --- Wed, April 9 4:30-7pm

Using a dataset on county demographics, presidential voting, and deaths of despair, this workshop covers how to use Stata to create regressions, test regression assumptions, manipulate data to make it satisfy regression assumptions, and understand model specification and interaction terms.

Week 12 *Manipulating and Analyzing Data in Stata (5th of the 7 workshop series)*

4/12-4/18 Location and registration by attending prior Stata workshops or instructor permission

Sun, April 12 4-6pm (Also offered in Week 10)

This workshop builds on the prior workshops in the 7 workshop series. It covers the basics of loops, the importance of documenting your work and version control, and creating codebooks for your data. We also touch on more advanced techniques and tricks for processing data, including fuzzy matching, looping through files, systematically renaming variables, regular expressions, and invisible characters.

Week 13 *Visualizing Data (7th of the 7 workshop series, or a standalone workshop)*

4/19-4/25 Location and registration required at <http://bit.ly/dataVizSpring2020>

Offered 2 times: Sun, April 19 4-6:30pm --- Mon, April 20 4:30-6:30pm

Data visualization is a powerful tool to communicate a point, but can also be misused. We cover how to create maps and basic descriptive images in R, and show how they can be manipulated to deceive.

Week 14

4/26-5/1

Reading 5/2-5/6

Finals 5/7-5/14

Week 1
Jan 22-26
Week 2

Jan 29-Feb
2

Week 3

Feb 5-Feb 9

Week 4

Feb 12-Feb
15

Week 5

Feb 19-Feb
23

Week 6

Feb 26-Mar
2

Week 7

Mar 5-Mar 9

break
Mar 10-Mar
18
Week 8

Mar 19-23

Week 9

**Mar 26-Mar
30**

Week 10

**April 2-
April 6**

Week 11

**April 9-
April 13**

Week 12

**April 16-
April 20**

Week 13

**April 23-
April 27**

Week 14

**April 30-
May 4**

**Reading
Week**

**May 5-May
9**

Introduction to Stata

This introductory session is based on a survey conducted on workshop participants. We cover the basic files, running commands, creating summary statistics and graphics, dealing with outliers, manipulating labelling variables. We also work through how to deal with errors (which are inevitable, but manageable, statistical package). This will be hands on, practical experience: each participant will be using the program in real time.

Introduction to Interpreting Regressions

Many books and papers present regression results to support their arguments. There are some simple ways to read these tables that let you quickly understand the main points and common pitfalls. We cover how to read regression tables and graphs efficiently, what the numbers for coefficients and statistical significance mean, and common ways in which regressions are misused.

Introduction to R

This introductory session is based on a survey conducted on workshop participants. We cover the basic files, running commands, creating summary statistics and graphics, dealing with outliers, manipulating data, creating data. We also work through how to use the package antitrust. We also work through how to deal with errors (which are inevitable, but manageable, in any statistical package). This will be hands on, practical experience: each participant will be using the program in real time.

Where to Find Data

Many research ideas are generated by creating questions answerable by existing data sources. We cover commonly used data sites, how to navigate these sources, and how to search for specific questions in surveys. We also use Google Sheets to easily create some basic graphics to summarize data on public opinions of government.

Introduction to ggplot in R

Ggplot is a powerful package devoted to creating beautiful graphics. We will cover how to plot regression lines, histograms, scatterplots, best fitting lines, and facet plots. I will use these plots to highlight how to interpret interaction terms.

Cherry Picking Data: The Widespread Problem of P-Hacking

P-hacking is a common way to cherry pick your results by selecting regressions and t-tests that are artificially significant. The kicker is that many people don't know they are doing it. We look at the connection, or lack of connection, between political parties and national economic health. We discuss how p-hacking happens: a common problem in social science research, and how to avoid it.

Game Theory in Social Science: When Philosophy Meets Math

Game theory is mathematical philosophy for social science. We start with the classic prisoners' dilemma. In prison terms we use a tastier incentive to reach the best outcome considering your opponent's best strategy. We discuss the ways that game theory has improved social science theory, and where it still needs improvement.

Best Practices in Data Analysis and Coding (Some Stata background required)

This workshop covers the tools required to produce easily interpretable, reproducible code. This includes cleaning practices, the importance of consistent output, how to structure a do file and comment smartly, naming conventions. Most importantly, we cover the importance of theoretical development for any data project, and how to concisely describe the purpose of your code as it relates to your theory.

Data Visualization: Evil Plots, Bad Plots, and Angelic Plots

Data visualization is a powerful tool to communicate a point, but can also be misused. We discuss the highlights in data interpretation, how graphics can highlight important theories, how to create your own graphics in Google Sheets, and how graphics are used to misinform.

No Workshop

No Workshop